MONT 107N – Understanding Randomness
Information and Sample Questions for Midterm Exam

General Information

The midterm exam will be given in class on Friday, March 19. It will cover the material from Chapters 16 - 21 of Freedman, Pisani, Purves. This means in particular:

1. Box models for chance processes and the Law of Averages. (Chapter 16)
2. Expected value and standard error (including the "shortcut formula" for a box with only two types of tickets). 0/1 boxes for classifying/counting. Using the normal curve to understand the likelihood of different sums of values on drawing from a box. (Chapter 17)
3. Probability histograms and the normal approximation as the number of draws gets large (the Central Limit Theorem) (Chapter 18)
4. Sampling concepts (including different potential sources of bias in a sample) (Chapter 19)
5. Chance errors in sampling – using a box model to understand the probabilities of different outcomes. Understand that the basic formulas we have used are essentially approximations in two different ways: First we use the formulas for sampling with replacement even though that is not strictly what is going on (but OK as long as the population is large relative to the size of the sample). Second, to estimate the SE, we use the observed proportion from the sample (not the actual population proportion, which is not known!) (Chapter 20)
6. Accuracy of percentages – EV and SE for percentages, using the normal curve, confidence intervals for percentages. (Chapter 21)

In general, be prepared for some questions that ask for calculations using these ideas *and also* for more conceptual questions about how and why they work and what they mean. See the practice exam questions below for examples of each type.

I will supply a copy of the normal curve area table from the text for your use on the exam.

No use of cell phones, pagers, I-pods, or any other electronic devices beyond a calculator will be allowed during the exam – please turn them off and stow them in your backpack.

Sample Exam Questions

Be sure to provide explanations for your answers if that is asked for. (*Note:* The actual exam will be somewhat shorter than this; there are parts here to illustrate most of the different kinds of questions that might appear.)

I. Assume 125 draws with replacement are made from the following box:

1 3 5 7

A) What are the EV and SE for the sum of numbers on the drawn tickets?
B) What are the chances that sum of the numbers will be 480 or larger?

II. Consider the following box.

0 1 2 2 2 3 3 4 4 4

A) If a single draw is made from the box, construct the probability histogram for the number that is drawn.
B) If 1000 draws with replacement were made from this box, draw an approximate probability histogram for the value of the sum. Show carefully the location of the EV and the SE for the sum relative to your histogram.
C) Now, assume 200 draws with replacement are made from this box. What box model would you use in order to model *the number of 4s* that are chosen?
D) Exactly 200 draws are made and 4s are drawn a total of 67 times. What is the difference between for the number of 4s for these 200 draws and the expected value of the number of 4s (that is, the *chance error* for this sample)?
E) Would you consider the chance error from part D to be a large chance error, roughly an average chance error, or a small chance error? Explain your answer. (Hint: You may want to do some additional calculations to decide on your answer.)

III. On March 3, 2010, the *New York Times* reported results of a survey in which researchers contacted 600 people who graduated from a variety of high schools across the country in the last 4 to 12 years. The study was sponsored by the Bill and Melinda Gates Foundation, which has sought to shed light on low completion rates at both the high school and college levels. For the purposes of this question, assume the 600 survey respondents were a simple random sample of all recent high school graduates. In this sample, 402 people indicated that they thought their high school guidance counselor was "poor" or "fair" in helping them decide what colleges or technical schools would be good for them.
A) Based on this sample, construct a 95% confidence interval for the the percentage of all recent high school graduates who think their guidance counselor was "poor" or "fair" in helping them decide what colleges or technical schools would be good for them.
B) The *Times* reported that the survey had a sampling error of $\pm 5\%$. Does that match your confidence interval? If not, can you suggest where the 5% figure comes from?

IV. A sociologist takes a simple random sample of size 300 from the 27000 students at a large state university. In the sample, $222/300 \times 100\% = 74\%$ are undergraduates. The SE for the percentage is calculated as 2.5% and the sociologist writes down the 95% confidence interval for the percentage as $74\% \pm 5\%$. For each part, say whether the statement is true or false, and explain:
A) From the information in this sample it is believable that the actual percentage of undergraduate students is 78%.
B) The range from 69% to 79% is a 95% confidence interval for the percentage of undergraduates in the sample.
C) The 95% figure comes from an area under the normal curve.

D) The probability that the true population percentage of undergraduates is in the interval 69% to 79% is 95%.

E) If a second sample of size 300 was taken in the same semester, and a 95% confidence interval was computed from that sample, then the result could be 67% to 79%.

F) If samples of size 1200 were taken in the same semester, then on average, we would expect the SE to decrease and the interval to be narrower.

G) The process of drawing the sample is like drawing 300 from a box with 27000 tickets and recording the sum, where the ticket for each first-year student has the value 1, the ticket for each second-year student has the value 2, the ticket for each third-year student has the value 3, the ticket for each fourth-year student has the value 4, and the ticket for each graduate student has the value 5.